MOBVIS FP6-511051
Deliverable Report

D.7.1 revised
“Software and report on mobile solutions for advanced vision tasks, partially available on the project website”

Vision Technologies and Intelligent Maps for Mobile Attentive Interfaces in Urban Scenarios
Project co-funded by the European Commission
Sixth Framework Programme (2002-2006)
Information Society Technologies
FP6-2002-IST-C / FET Open
STREP

Due date of deliverable:	April 30, 2007 (month 24)
Actual submission date:	December 4, 2007
Start date of project:	May 1, 2005
Duration:	36 months

Work package	**7 – Demonstrator and Dissemination**
Task	**1 – Mobile Vision Technology**
Lead contractor for this deliverable	**JRS**
Editor	**Lucas Paletta**
Authors	**Lucas Paletta, Andreas Wimmer, Mårten Björkman, Jan-Olof Eklundh, Katrin Amlacher**
Quality reviewer	**UL**

Project co-funded by the European Commission within the Sixth Framework Programme (2002–2006)		
Dissemination Level		
PU	Public	X
PP	Restricted to other programme participants (including the Commission Services)	
RE	Restricted to a group specified by the consortium (including the Commission Services)	
CO	Confidential, only for members of the consortium (including the Commission Services)	

CONTENT

1. EXECUTIVE SUMMARY

This deliverable describes the preliminary status of work within Task 7.1 (Mobile Vision Technology) until project month 24. The goal of this task in general is to provide mobile, server independent algorithms for advanced vision tasks that enable on-site recognition of objects of interest on the mobile device. In order to achieve this objective, it is important to investigate general performance bottlenecks for mobile object recognition tasks and accordingly analyse cost functions that would support any decision making in a scheduling that would distribute processing either to client or server based computing. Investigations of this kind would be focused on state-of-the-art technology, in particular, regarding camera equipped phones and PDAs.

This deliverable is supposed to represent a living document and will be updated until the official end of Task 7.1, i.e., until project month 36.

2. INTRODUCTION

2.1 MOTIVATION

This deliverable describes the preliminary status of work within Task 7.1 (Mobile Vision Technology) until project month 24. The goal of this task is to provide mobile, server independent algorithms for advanced vision tasks (see Tasks 3.2-3.4) that enable on-site recognition of objects of interest on the mobile device. In order to achieve this objective, it is important to investigate general performance bottlenecks for mobile object recognition tasks and accordingly analyse cost functions that would support any decision making in a scheduling that would distribute processing either to client or server based computing. Investigations of this kind would be focused on state-of-the-art technology, in particular, regarding camera equipped phones and PDAs.

This deliverable is supposed to represent a living document and will be updated until the official end of Task 7.1, and the project itself., i.e., project month 36. In this first, preliminary version, state-of-the-art is presented and analysed with respect to promising directions of work in the near future. In addition, we are describing the working environment that is used for mobile vision technology in MOBVIS, i.e., the image analysis software framework IMPACT, and the Mobile Windows 2005 emulator software package and the PDA Fujitsu/Siemens Pocket Loox.

Finally, this deliverable includes a concrete outlook on work in MOBVIS' third project year.

2.2 STATE-OF-THE-ART

Mobile computer vision - in terms of simple tracking tasks for outdoor augmented reality in urban scenarios - was first considered in [FMH97]. [BEH99] considered vision based positioning with a system based on the extraction of silhouettes for precise orientation of camera and user's view. The system was intended for use in natural terrain environments. Typical mobile vision approaches intend to extract discrete features, such as, to geo-reference lines in image sequences [CSH99], artificial landmarks [RIB02], or silhouette based information [BHK02]. Coors et al. [CHK00] match 3D information from on-journey images with the corresponding information from a virtual city model. While these techniques proved potential applicability giving initial results, serious shortcomings were identified in the time requirements of the used image analysis techniques.

Worldwide, a few projects have been performing massive attempts to focus on object type recognition from PDA based vision technology: (i) INTERACT (Interactive Systems Lab at Carnegie Mellon University, Pittsburgh, PA [INT03]), investigating fundamental vision interfaces for PDA based interaction, such as PDA-based indoors face recognition [YAN02], and initial results on outdoors text translation [ZHA02], (ii) AR-PDA (Siemens, Paderborn, Germany [ARP04]), working on augmentation of PDA based imaging of industrial products with manual information (iii) MARS (Computer Graphics and User Interface Lab, Columbia University, NY), introducing 3D computer vision for tracking in urban environments [MAR04], and (iv) INMOVE [INM04], an ongoing EU funded project coordinated by the VTT Multimedia Group, Finland, that introduces intelligence in video based applications for mobile users.

More recent engagement in the direction of mobile vision technology and its context in the frame of research and industrial development is described as follows.

[TYD05] used images of objects as queries is a new approach to search for information on the web. Image-based information retrieval goes beyond only matching images, as information in other modalities also can be extracted from data collections using image search. They demonstrated a new system that uses images to search for web-based information, introducing a point-by-photograph paradigm, where users can specify an object simply by taking pictures. Their technique uses content-based image retrieval methods to search the web or other databases for matching images and their source pages to find relevant location-based information. They developed a prototype on a camera phone and conducted user studies to demonstrate the efficacy of our approach compared to other alternatives.

Rohs and Zweifel [RZ05] evaluated the feasibility of using camera equipped mobile phones to act as sensors for 2-dimensional visual codes. The codes can be attached to physical objects and act as a key to access object-related information and functionality. The use of mobile phones is interesting in this scenario, because mobile phones are ubiquitously available devices providing constant wireless connectivity, and models with integrated cameras are getting more and more popular. Using the integrated camera as a sensor thus offers a natural way of detecting objects in the user's immediate surroundings.

TinyMotion [WZ06] is a software approach that detects the movements of cell phones in real time by analysing image sequences captured by the built-in camera. Typical movements that TinyMotion detects include - horizontal and vertical movements, rotational movements and tilt movements. As a result, a user can activate and access different functions of a phone (for example, scrolling and selecting the phone menu, Zooming in/out pictures, Moving an on-screen cursor to a given region and even gesture/handwriting input) by moving, tilting or rotating the phone. Different from existing research work, TinyMotion does not require additional sensors (accelerometers, motion sensors), markers (barcodes, anchor symbols, dots) or external computing powers and can run on today’s main-stream camera phones without hardware modifications. Even more, TinyMotion can detect camera movement reliablely under diverse background and illumination conditions.

[RSF06] use a camera equipped mobile phone for indoor localisation. In this system, the smart phone is worn by the user as a pendant and images are periodically captured and transmitted over GPRS to a web server. The web server returns the location of the user by comparing the received images with images stored in a database. They tested the system inside the Computer Science department building, preliminary results show that user’s location can be determined correctly with more than 80% probability of success. As opposed to earlier solutions for indoor localization, this approach does not have any infrastructure requirements, and the only cost is that of building an image database.

[JFX06] from Microsoft Research Asia designed and implemented a system named Photo-to-Search to carry out queries from camera phones simply by taking some photos of interested objects. The captured pictures are compared with a large amount of Web images to select the ones which contain the same prominent object. Consequently, the related information is extracted from the Web pages where the matched images locate. Data of large buildings, storefronts and products were collected and corresponding kinds of queries were specifically demonstrated to show the efficiency and the effectiveness of their system, announcing that image databases with more than 6000 images could be queried under satisfying constraints.

D2 Communications and Bandai Networks [D2C06] cooperated in in Japan to make a mobile visual searvch service called ‘Search by Camera! ER Search’ available. The service makes use of a mobile phone’s camera and it has first been introduced in NTT DoCoMo’s N902iS phone. Search by Camera! ER Search’ is a mobile phone service based on image recognition technology and i-Appli, enabling the users of built-in camera phones to purchase products and get product information by making use of mobile phone’s camera. When the user takes pictures of wine labels, shopping catalogues, CD covers, etc., and sends the images to the server, it compares the images with pre-installed data and gives the user various information on the product. ‘Search by Camera! ER Search’ uses Evolution Robotics’ ViPR (visual pattern recognition) technology developed for mobile phones by Bandai

Networks in Japan. The service opens new marketing opportunities to companies as they can provide more visually attractive advertising via mobile phones that until now has mainly been limited to SMS.

Geovector and cybermap Japan enhanced the world’s first pointing based local search solution for mobile phones, in terms of its ‘Mapion Pointing Application’ [GEO06]. While this approach does not make use of image analysis, it uses GPS, a digital compass and a digital city map to offer users a 3D view on a currently hypothesised environment, enabling the user to select 3D objects and connect to annotating information in response. It offers new mobile local search capabilities including first of its kind user driven, opt-in advertising, sponsored categories and preferred placement. This intuitive pointing interface provides access to information on 700,000 Points of Interest across Japan that are now available on over 2 million Sony Ericsson , Kyocera and Casio mobile phones using the KDDI network.

Mobile Vision Technologies [MVT07] is a German small medium enterprise which is declared to be a follow-up company of Evolutionary Robotics (CA; USA). It developed a technology, called Visual Search on Demand (VSoD) which applies visual object recognition on mobile imagery. They propose target applications, such as, face recognition (Look-Alike - finding famous faces and a matching score with respect to any face image fed into the system), or a mobile marketing tool that identifies company brands or associated pages in newspapers in order to link to related web pages.

Related research on vision based object detection and recognition has recently focused on the development of local interest operators [KAD01,MIK02] and the integration of local information into occlusion tolerant recognition [WEB00,FER03]. While concern has been taken specifically with respect to issues of scale invariance [FER03], wide baseline stereo matching performance [MIK02,OBD02], or unsupervised learning of object categories [WEB00,AGA02,FER03], the application of interest point operators has not yet been investigated about the appropriate cost function that should evaluate performance with respect to both the information content and the robustness of extraction, e.g., of objects of relevance in urban scenarios.

While first steps towards informative features [FPB04] and illumination insensitive recognition [BWL04] have already been initiated by MOBVIS consortium members, MOBVIS intends to pursue this investigation with the goal to arrive at a mobile object recognition demonstrator. Characteristic approaches in MOBVIS towards enabling mobile vision services on large image databases are in the direction of geo-contextual object search [PF07], context based search for object detection [PL07], and, innovative methodologies for matching in high-dimensional feature spaces for finding most informative features [OL07]. In regard of contextual image search, the development towards an Attentive Interface that takes advantage of multimodal information for cutting down search in image feature space is – at the time of the generation of this document - in its preliminary stages.

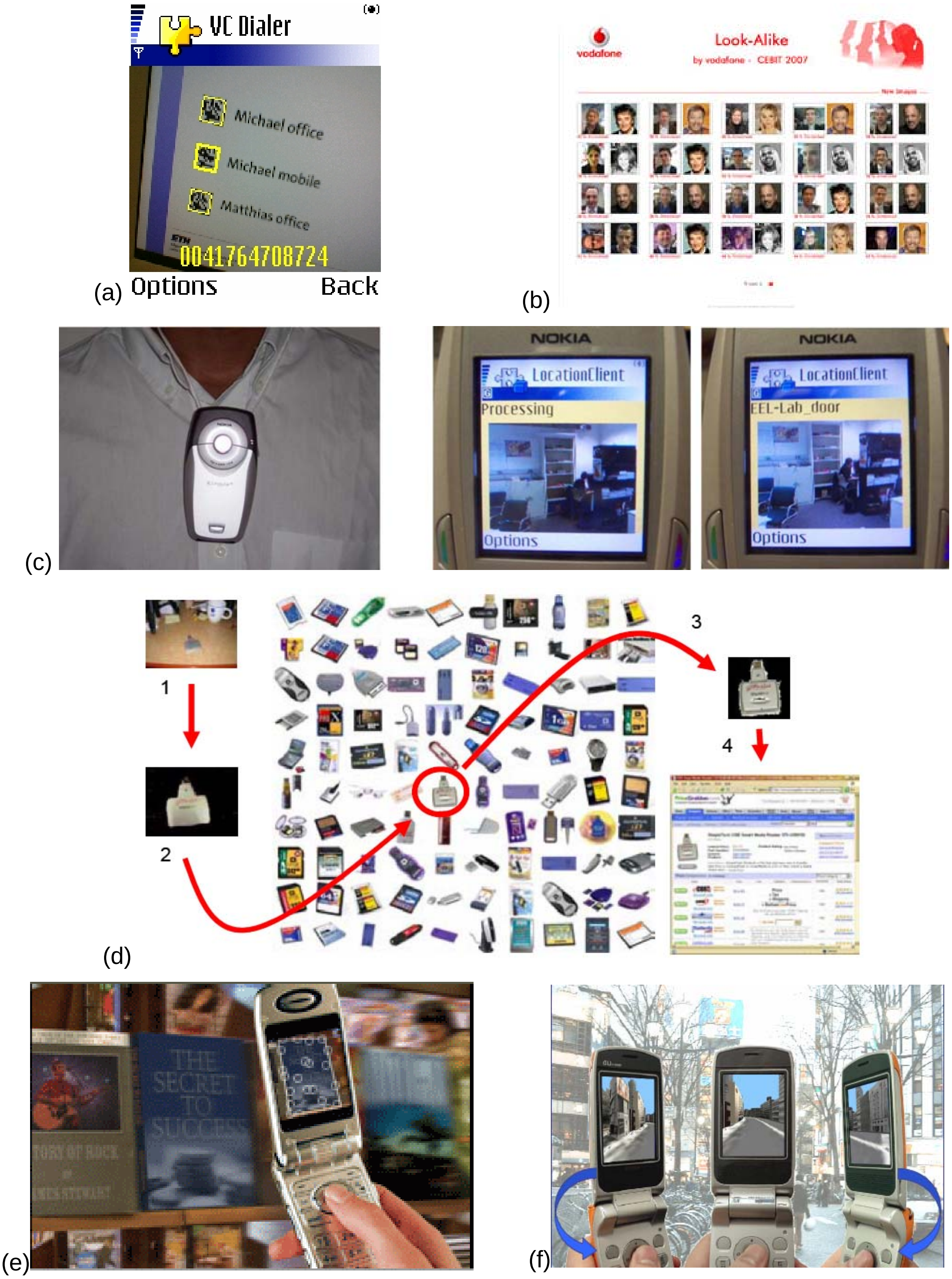

Figure 1. (a) [RZ05] evaluated the feasibility of using camera equipped mobile phones to act as sensors for 2-dimensional visual codes. (b) [MVT07] developed a visual search tool to match faces on mobile imagery with famous ones. (c) [RSF06] use a camera equipped mobile phone for indoor localisation. (d) [TYD05] used object images as queries to search for information on the web. (e) [D2C06] made mobile visual search service available for more than 2 million phones in Japan. (f) Geovector enhanced the world's first pointing based local search solution for mobiles [GEO06].

2.3 MOBVIS APPROACH TO MOBILE VISION TECHNOLOGY

In MOBVIS, we are primarily interested in identifying general challenges in computer vision in the context of mobile vision services. According to this approach to mobile vision, it is not of prior interest to build a complete client based computer vision system, but, instead, to select a viable approach and then to analyse it from a general viewpoint, i.e., identifying bottlenecks with respect to classical computer vision, propose cost functions that would support decison making about whether to use client or server based image analysis, and outline a research roadmap that would point into promising research directions in order to target at reasonable future implementations of mobile vision technology.

As a first step into the targeted direction, we firstly identified relevant directions in the state-of-the-art in mobile vision services (Section 2.2). Secondly, we selected an existing sofwtare framework for image analysis, i.e., the IMPACT framework (Section 3.3) for testing a characteristic mobile vision service as it is developed in MOBVIS, i.e., object recognition on mobile imagery (see [PF07]). In the preliminary stage of this Task, we ported most relevant image processing functions into an emulator software for Microsoft Windows Mobile 2005, a software which is highly promising to become a future standard in mobile devices (smartphones, PDAs, etc.), and relatively easy to work with. The next step in this work will be a thorough analysis of the ported software in terms of response times, workload and the capability to be improved in order to become a basis for a truly mobile vision service. We expect as final outcomes of this study,

- A description of cost functions for the evaluation of mobile vision software,
- A benchmark test for performance evaluation,
- An evaluation report on the mobile IMPACT software for object recognition in urban environments,
- A demonstrator on mobile visual object recognition, and
- A list of most promising applications in the context of mobile object recognition,
- A report summarising all results described above being published on the MOBVIS homepage.

The following Section on the impleplementation of mobile vision starts with general remarks on client- versus server-based computing of mobile imagery, describes then toolkits for the development of mobile vision software, outlines an overview on the IMPACT software framework, and finally provides an experience report on the porting of IMPACT software from a server based solution to mobile devices for the purpose of mobile vision.

3. IMPLEMENTATION OF MOBILE VISION

3.1 GENERAL IMPLEMENTATION ASPECTS IN MOBILE VISION

Mobile devices have general limitations compared to computers, which of course also have an impact on attempts to implement advanced computer vision algorithms on them. Key limitations of this type are due to power consumption and memory requirements. It is worth observing that most portable devices lack virtual memory handling, something that enforces an old-fashioned style of programming with explicit management of memory.

Other kinds of generic limitations concern optics and time-varying imagery. Advanced vision algorithms are usually developed and tested on high-quality images or video streams. Mobile devices often generate images of limited quality, which makes it difficult to port some of the more sophisticated methods directly to mobile devices. They need to be adapted to lower quality images, something that can be difficult. Video streaming is today generally not possible on mobile phones at least not in high resolutions. Hence, algorithms benefitting from temporally varying imagery, e.g. for simple tasks as noise reduction and obtaining multiple views for recognition or depth recovery, are not directly useful.

Until only a few years ago there was a high variability between mobile devices as computing tools. Although standardization is not on-going for them one can however observe a clear tendency of convergence when it comes to various aspects of computing and data acquisition. Hence, the lack of standards may not be a major problem in the years to come.

There is a considerable difference between both programming and using a mobile device compared to a standard computer. The difference lies in how interaction can take place. On devices such as a mobile phone one can use either keys or a pens, that is similarly to a mouse. In that respect there isn't much difference. However, the screen size on mobile devices is usually very limited and don't allow any advanced handling of windows. Hence, the both the programmer and the user have a rather limited view of what is going on, which puts strong demands on the interaction. Whether such kinds of limitations will be remedied or not in a reasonable future or not is an open question. Interaction problems of this nature are widely studied in the HCI community.

Irrespective of the optics mobile devices are today sometimes provided with cameras giving images as large as and with the resolution of those obtained by digital cameras. Transferring such images by http or mms in general takes more than 10 seconds. Hence, if images are transferred to servers the turn-around time can be considerable and users may experience this as inconvenient. However, these limitations are not technically motivated. Instead they depend on priorities made by e.g. operators and the infrastructure. Today tasks such as on-line TV and downloading of movies seem to be in focus. In a longer perspective it is likely that all kinds of image transfer will be fast. As a consequence one can foresee that much of the advanced computer vision processes will take place on servers rather than on clients.

In general one can predict that power consumption will continue to put limitations to what can be done on mobile devices. The interaction that both end users and programmers need to perform need to be developed substantially and it is expected that this will be done by the HCI community. A remark worth doing is that processing anything on a client like a mobile phone usually requires downloading software and providing a number of settings. Such operations will probably always constitute obstacles to the use of advanced software on mobile devices, at least for the everyday user.

3.2 WORKING ENVIRONMENT FOR MOBILE IMAGE ANALYSIS

In order to test mobile vision technology in MOBVIS, we decided to decide for a working environment that would encompass features of a state-of-the-art hardware together with making use of an existing image analysis framework that will be applicable thereon.

Currently, there is a large market for camera equipped mobile phones. Moreover, mobile phones, in particular, the category of smartphones, are tending to overtake the role of PDAs, laptops and UMPCs (ultra mobile PCs) by increasingly including PC features while at the same time fully offering classical communication features and the usability of mobile phones.

The choice of experimentation platform is a smartphone that is using the latest edition of the Microsoft Windows Mobile operating system, which is commonly seen as the emerging OS in mobile computing platforms. In particular, the Fujitsu Siemens Pocket LOOX, actually a kind of PDA, is combining features from mobile phones in terms of its connectivity, together with camera sensor availability and substantial processing capabilities (416 MHz CPU ; see Figure 2, Figure 3).

Processor:	Intel® PXA270 416 MHz based on Intel® XScale™ microarchitecture
Memory:	64 MB System memory (RAM) 128 MB Flash memory (ROM)
Weight:	ca. 195g
Dimensions:	126 x 64 x 21 mm
Display:	2.4-inch, 240 x 240 color transflective TFT touchscreen, 65.536 colors, LED backlight
Transceiver & Power Amplifier:	GSM900/1800/1900MHz & UMTS 2100MHz operations
Battery:	Exchangeable Lithium-Ion battery: 150 hours standby 4 hours talk time
Operating System:	Microsoft® Windows Mobile™ 5.0 Phone Edition
Integrated:	GPS (SiRF GSC3F), Wireless LAN (802.11b/g, Wi-Fi certified), Bluetooth 2.0
Camera:	VGA camera (640x480) for video telephony via UMTS, 2 Megapixel auto focus camera (1600x1200) for picture capturing
Interfaces:	1 x built-in microphone 1 x speaker, 1 x receiver for VoIP 1 x headphone (3.5 mm) 4 pin USB 1.1 (slave) via sync cable USB 1.1 (host) via sync cable

Figure 2. Technical data of the choice of mobile computing platform, the Fujitsu Siemens Pocket LOOX T830.

Figure 3. Appearance of the Fujitsu Siemens Pocket LOOX T830 hardware.

While developing the mobile front end for the MOBVIS system, thorough testing of the application is necessary. This is done by deploying it on both the mobile device as well as using the Windows Mobile 2005 emulator.

The emulator provides a runtime environment identical to any PDA and Windows Mobile smartphone on the market, so it is possible to run the front end on a wide spectrum of different configurations such as display resolution, CPU frequency and / or available memory. The emulator incorporates the full capability of the .NET compact framework, which is composed of communications (Bluetooth, WLAN, UMTS and GPRS), user interface controls and handling of multimedia data (image / sound / video capturing and processing).

In addition to this, in combination with the Visual Studio 2005 IDE, the emulator suite offers various debugging mechanisms, which render the developing process much more efficient. The debugger includes features like conditional breakpoints with hit counts, breaking on exceptions with a detailed description of the exit status, stepping-through code, viewing the current state of objects and enhanced visualization of complex types.

Finally, besides the benefit in application development, the emulator is best suited for presenting the application on a large scale screen.

3.3 THE SOFTWARE FRAMEWORK IMPACT

At the Institute for Digital Image Processing most of the processing tasks which involve higher level image analysis capabilities are performed within the IMPACT software framework. IMPACT[1] (Image Processing and Classification Toolkit) has been designed from the beginning to be a framework which supports the expert in his goal to develop complex processing chains from individual standard image processing components which ease the use and which enhance the applicability of imagery of any size (mobile imagery, video frames, remote sensed imagery) for a larger audience.

One of the primary goals of IMPACT has been the use of a class based programming language (C++) and the utilisation of this language to reduce the development effort to a high degree. Hence the products and programs built within the IMPACT framework are highly modular and reusable. This modularisation of reusable components plays the key role when creating customised processing chains for customers, since the development effort is considerably reduced.

The IMPACT framework consists of several components which are developed to cover specific tasks and simultaneously provide modularised parts for other research and development areas. The most relevant components (among others) are listed below:

- Basic image manipulation filter library (scaling, arithmetic processing, …)
- Statistical analyses library (regressions, histogram matching, …)
- Noise manipulation library (speckle reduction, …)
- Library for object based analyses (interest points, classification, …)
- Visualization and statistical analyses of small to large imagery (ImpactViewer)
- Generic image handling facilities due to indexing capabilities (multi-gigabyte files are no problem)

IMPACT includes many standard image processing functions, such as, generic algorithms (convolution, Canny edge detection, edge coherence, Roberts/Sobel/Prewitt operators, resampling, blob detection, etc.), segmentation methods (region growing, k-means, fuzzy k-means, neural gas segmentation, region competition, etc.), texture analysis (wavelet transform, local information transform, law filter, etc.), and statistics based methods (cluster analysis, PCA, BFS feature selection, genetic algorithms, neural networks, etc.). In MOBVIS, we have used functionalities, such as for example, wavelet transforms, decision trees, learning of cascaded classifiers, SIFT descriptors, and segmentation facilities.

The conception of the software package IMPACT facilitates the development and implementation of customised processing chains. Software modules can be arbitrarily linked to end-to-end processing lines, or individual modules can be easily integrated into already existing processing environments.

[1] © JOANNEUM RESEARCH Forschungsgesellschaft mbH, Institute of Digital Image Processing

3.4 PORTING IMPACT TO MOBILE TECHNOLOGY

The porting of routines for image analysis from the software framework IMPACT to a mobile platform proved to be a relatively difficult task. Until recently (prior to the release of Windows Mobile 5) the issue would not be solvable at all. IMPACT is a sophisticated and advanced image processing framework which has been build using C++ as programming language, more precisely IMPACT is makes use of most of the modern C++ language features like template programming, exception handling operator overloading, etc. The compilers available before the release of Windows Mobile 5 where not capable of providing these features, hence porting IMPACT at that time would essentially have meant a new implementation from scratch.

With the release of Windows Mobile 5 a new C++ compiler is available which provides all the necessary compiler features. Therefore the porting of IMPACT was envisaged and some initial investigations proved that a porting had good chances to be successful after all. The remaining problems where essentially that the Windows Mobile 5 platform does not provide all the Windows API's which where used by IMPACT or one of its external components.

In a first step it was necessary that profoundly required external libraries where ported to the Windows Mobile 5 platform. For the GDAL library this was a relatively easy task as it already provided some support for the mobile platform. Using this support it was possible to offer the possibility to read or write tiff file on the Windows Mobile 5 platform. The second step was a bit more difficult, as the BOOST package did not provide support for the mobile platform. After some modifications it was finally possible to compile at least the absolutely required boost-thread library, all other used boost libraries would have required much more work as they where making heavy use of some API functions and structures which where not available on the mobile platform (most notably some functions from the Windows posix subsystem).

After the porting of these external libraries work started to port the IMPACT framework itself. As the development of IMPACT started on the Linux platform, there where quite a lot of places where IMPACT was using the POSIX subsystem on the WIN32 platform. The biggest problem where the *fileio* interfaces, due to the fact that the traditional open/read/write functions had to be replaced by the Windows style CreateFile/ReadFile/WriteFile functions.

To avoid a fork of the IMPACT code base, a new platform independent *fileio* library had to be designed and implemented. After the completion of this library the mainline IMPACT development has been ported to this new library, and proved to be a reliable replacement for the original code.

The porting of the QT package, which is used for the graphical user interface, was assessed as requiring to much work or even being impossible at all. Hence the ported algorithms had to be used via command-line without possibility of any user interaction. This issue was considered of secondary interest as a proper Windows Mobile application would need a specifically designed user interface anyway, and in this case IMACT would serve as image processing library used by this application.

Another source of problems for the porting were the difficulties in compiling for the Windows Mobile 5 platform. To be able to produce a proper executable or library it is necessary to specify a number of defines, to set a number of compiler options and to provide the proper entry point for a program. An additional source of challenges is the fact that a Windows Mobile Application only gives some kind of generic error message if a program could not be started, there is no indication which DLL´s are missing or that the program has been compiled for the wrong architecture.

Once all these problems where solved the next step is the evaluation and analysis of some specific algorithms on a typical mobile device in comparison to a normal Windows PC system.

4. EVALUATION OF MOBILE VISION ALGORITHMS

4.1 EVALUATION IN THE CONTEXT OF OBJECT RECOGNITION

We intend to investigate the performance of mobile vision services in the context of object recognition. The following components can be identified for any object recognition framework,

- Image feature analysis,
- Image segmentation,
- Matching of analysed with reference object features,
- Object hypothesis generation ,
- Decision making about object identity (or/and category).

In Task 7.1, we intend to analyse the performance of client- and server-based computing with respect to all these components. In the frame of mobile object recognition proposed in [PF07], we will specifically analyse the following functions, (i) SIFT descriptor extraction from the raw mobile image, (ii) nearest-neighbour search with respect to reference descriptors, (iii) image segmentation with respect to hypothesised descriptor labels, and (iv) decision making with respect to object hypotheses that are associated to the extracted image segments.

It is anticipated that the components with most workload on real-time processing will be identified as being components (i) and (ii), respectively. (i) SIFT descriptor extraction relies on a multi-scale approach for image representation, involving Gaussian filtering on a hierarchy of image scales. Most probably, we will select a specific, most probable scale for processing and finally will have to accept some degradation in quality, ie.., accuracy in the recognition, while being sufficiently satisfying in the response time. (ii) Nearest-neighbour search requests considerable resources in storage, and also in processing time, therefore we will also experiment on various approaches to compromise between accuracy and time response in the application. However, the detailed evaluation of the mobile image processing will be performed indetail in the third project year, and documented in the final update of the Deliverable D7.1 report (due until project month 36).

4.2 WORK PLAN FOR OBJECT RECOGNITION WITH MOBILE VISION TECHNOLOGY

In the following, we outline a concrete plan for the implementation and evaluation of MOBVIS technology on a mobile device. As outlined in Sec. 3.2, a smartphone, but also the hardware used for the MOBVIS final demonstrator, a Vaio Sony UMPC UX-1 will serve as platform for the implementation and the experiments.

The key interest in MOBVIS is to evaluate Attentive Interface technology – in the context of mobile vision technology, we will evaluate the performance on the mobile device.

Evaluation will be on *three major functionalities*:

- **Geo-indexed object recognition.** This functionality makes use of Attentive Interface components, in particular, using the geo-services for a selection of object hypotheses.

- **Logo detection and recognition.** This functionality requires more vision components than the geo-indexed object recognition since it requires more sophisticated segmentation methods.
- **Street name based positioning.** Here again, multiple vision components will be under evaluation, such as segmentation and an OCR component.

Benchmark tests will be performed in the scenario of the MOBVIS final demonstrator, i.e., in the Inner City of Graz. Geo-referenced image databases have already been acquired during the beginning of the third project year (GUIS107, Logo image database).

For *performance evaluation*, we will outline a detailed overview of involved functional components, control and data flow, and associated bandwidth and processing times with the given hard- and software. A *user group* will be involved in order to evaluate the usability of the choice of approach.

In the case of *geo-indexed object recognition*, we will apply the geo-services in *attentive mode*, i.e., feeding the priors gained from the geo-context into the visual object recognition module in order to speed up the processing time and reduce required resources.

In the case of *Logo detection*, logos will not only be detected at unique geographic locations, but also regardless to the location and building providing the visual background.

The results of the implementations, the benchmark test, and the performance evaluation will be described in a *final update of Deliverable* D7.1.

The targeted functionalities will be available as *part of the final demonstrator*, i.e., implemented on the mobile device and ready for reviewing experiments at the final review meeting.

4.3 OVERALL GUIDELINES FOR FUTURE WORK ON MOBILE VISION

The results of the initial evaluation show that there is still quite a huge performance hit which has to be expected when running programs on a mobile device. In future it is expected that the performance on these platforms will increase, however it seems unlikely that the gap to a typical PC type hardware will be closed.

To improve the situation on such mobile devices it is necessary to carefully design, adjust and optimize applications which are expected to run on these platforms. It seems highly unlikely that it will be possible in the near future the simply recompile and run a PC type application on a mobile device. Applications for a mobile device will continue to require careful analysis and realisation to have a chance to be a successful application.

5. SUMMARY AND OUTLOOK

5.1 SUMMARY

This report described the initial steps for the investigation of mobile vision technology in the context of MOBVIS. On the basis of a description of the state-of-the-art in mobile vision services, we intend to focus research on the general capabilities of image analysis software to be appropriate for the application in mobile devices. MOBVIS started in this direction by using a standard emulator software for the porting of the image analysis software framework IMPACT onto Microsoft Windows Mobile Phone/PDA 2005.

5.2 OUTLOOK

As described above, in the third project year, the evaluation of the IMPACT software with respect to mobile vision functionality and performance will be performed in detail.

- A description of cost functions for the evaluation of mobile vision software,
- A benchmark test for performance evaluation,
- An evaluation report on the mobile IMPACT software for object recognition in urban environments,
- A demonstrator on mobile visual object recognition, and
- A list of most promising applications in the context of mobile object recognition,
- A report summarising all results described above being published on the MOBVIS homepage.

The immediate next step in the evaluation will be the transfer of emulation software onto the mobile device and the profiling of software processing performance.

6. REFERENCES

[AGA02] Agarwal, S., Roth, D.: Learning a sparse representation for object detection, Proc. *European Conference on Computer Vision*, Vol. 4, pp. 113-130, 2002.

[ARP04] AR-PDA: A digital mobile Assistent for VR/AR-Content, http://www.ar-pda.de/, Paderborn, Germany, 2004.

[BEH99] Behringer, R., Registration for outdoor augmented reality applications using computer vision techniques and hybrid sensors, *Proc. VR ’99*, pp. 244-251, 1999.

[BHK02] Böhm, J., Haala, N., Kapusy, P.: Automated appearance-based building detection in terrestrial images. *International Archives on Photogrammetry and Remote Sensing IAPRS*, Volume XXXIV, Part 5, pages 491-495, ISPRS Commission V Symposium, Corfu, September 2002.

[BWL04] Bischof, H., Wildenauer, H., Leonardis, A.: Illumination insensitive recognition using eigenspaces. *Computer Vision and Image Understanding*, 2004.

[CHK00] Coors, V.; Huch, T.; and Kretschmer, U., Matching buildings: Pose estimation in an urban environment, *Proc. IEEE and ACM International Symposium on Augmented Reality*, pp. 89-92, 2000.

[CSH99] Chen, T., and Shibasaki, R., A Versatile AR Type 3D Mobile GIS Based on Image Navigation Technology*, Proc. IEEE International Conference on Systems, Man, and Cybernetics*, 1999, pp. 1070-1075.

[D2C06] http://www.gearlog.com/2006/05/11/index.php

[FER03] Fergus, R., Perona, P., Zisserman, A.: Object Class Recognition by Unsupervised Scale-Invariant Learning, *Proc. Conference on Computer Vision and Pattern Recognition*, Volume II, pp. 264-271, Madison, Wisconsin, 2003

[FMH97] Feiner, S., MacIntyre, B., and Höllerer, T., A Touring Machine: Prototyping 3D Mobile Augmented Reality Systems for Exploring the Urban Environment, Proc*. IEEE International Symposium on Wearable Computers*, Boston, MA, 1997, pp. 74-81.

[FPB04] Fritz, G., Paletta, L., and Bischof, H., Object Recognition using Local Information Content, Proc. IAPR *International Conference on Pattern Recognition*, ICPR 2004, Cambridge, UK, August 22-26 2004, in print.

[GEO06] http://www.geovector.com/press/netdimensions.html

[INM04] VTT Multimedia Group, http://www.vtt.fi/multimedia/index.html, Espoo, Finland, 2004.

[INT03] Interactive Systems Labs, Carnegie Mellon University, Pittsburgh, PA, http://www.is.cs.cmu.edu/papers/.

[JFX06] Jia, M., Fan, X., Xie, X., Li, M., Photo-to-Search: Using Camera Phones to Inquire of the Surrounding World, *Proc. 7th International Conference on Mobile Data Management* (MDM'06), p. 46.

[KAD01] Kadir, T., Brady, M.: Scale, Saliency and Image Description. *International Journal of Computer Vision.* 45 (2):83-105, 2001.

[MAR04] Feiner, S., Höllerer, T., Gagas, E., Hallaway, D., Terauchi, T., Güven, S., MacIntyre, B., Computer Graphics and User Interfaces Lab, Columbia University, http://www1.cs.columbia.edu/graphics/projects/mars/mars.html, New York, USA, 2004.

[MIK02] Mikolajczyk , K., Schmid, C.: An affine invariant interest point detector, *Proc. European Conference on Computer Vision*, vol. 1, 128--142, 2002.

[MVT07] http://www.mobile-vision-technologies.com/

[OBD02] Obdrzalek, S., J. Matas, J.: Object recognition using local affine frames on distinguished regions, Proc. *British Machine Vision Conference*, pp. 113-122, 2002.

[OL07] Omercevic, D., and Leonardis, A., High-dimensional feature matching: Employing the concept of meaningful nearest neighbors, *submitted to International Conference on Computer Vision*, ICCV 2007, October 2007.

[PF07] Paletta, L., and Fritz, G., Visual Object Recognition in the Context of Mobile Vision Services, *Proc. 5th International Symposium on Mobile Mapping Technology* (CD-ROM), May 28-31, 2007, Padova, Italy.

[PL07] Perko, R., and Leonardis, A., Context driven Focus of Attention for Object Detection, submitted to Paletta, L., and Rome, E., Eds., *Attention in Cognitive Systems*, Lecture Notes in Computer Science – LNAI, to appear October 2007.

[RIB02] Ribo, M., Ganster, H., Brandner, M., Lang, P., Stock, C., Pinz. A.: Hybrid tracking for outdoor AR applications. *IEEE Computer Graphics and Applications Magazine*, 22(6): 54-63, 2002.

[RSF06] Ravi, N., Shankar, P., Frankel, A., Elgammal, A., and Iftode, L., Indoor Localization Using Camera Phones.. *Proc. 7th IEEE Workshop on Mobile Computing Systems and Applications*, April 2006.

[RZ05] Rohs, M., and Zweifel, P., A Conceptual Framework for Camera Phone-based Interaction Techniques, *Proc. Pervasive Computing: Third International Conference*, PERVASIVE 2005, Lecture Notes in Computer Science (LNCS) No. 3468, Springer-Verlag, Munich, Germany, May 8-13, 2005.

[TYD05] Tollmar, K., Yeh, T., and Darrell, T. , IDiexis: Mobile image-based search on world wide web- a picture is worth a thousand keywords, *Proc. of Mobisys,* 2006.

[WEB00] Weber, M., Welling, M., Perona. P.: Unsupervised learning of models for recognition. Proc. *6th European Conference on Computer Vision*, Dublin, Ireland, 2000.

[WZ06] Wang, J., and Zhai, S., Camera phone based motion sensing: interaction techniques, applications and performance study, *Proc. 19th ACM symposium on User interface software and technology*, pp. 101-110, 2006.

[YAN02] Yang, J., Chen, X., Kunz, W.: A PDA-based Face Recognition System, *Proc. Sixth IEEE Workshop on Applications of Computer Vision*, Orlando, Florida, 2002.

[ZHA02] J. Zhang, X. Chen, A. Hanneman , J. Yang , Alex Waibel, "A robust approach for recognition of text embedded in natural scenes," *Proceedings of ICPR 2002*, Quebec City, August, 2002